

WWW 20 år – og stadig mest tekst

af Niels Brügger, Centerleder, Center for Internettforskning, Aarhus Universitet

Den 6. august 2011 var det præcis 20 år siden, www kom til verden. Det skete, da den engelske computerspecialist Tim Berners-Lee sendte sin første korte præsentation af WorldWideWeb til nyhedsgruppen alt.hypertext. Hermed var grunden lagt til, at internettet kunne få stor folkelig udbredelse - og det fik det. WWW og internettet bruges i dag til så mange gøremål, at vi ville have svært ved at vende tilbage til en periode uden www: offentlige myndigheder kommunikerer i stigende grad med borgerne via www, både små og store virksomheder har i dag et website, den offentlige debat og civilsamfundets aktiviteter foregår i stigende grad på www, ligesom også stadig flere af vores private sociale aktiviteter gør.

WWW nåede til Danmark i slutningen af 1992 og blev i de følgende år mere og mere udbredt, men hvordan har det danske web egentlig set ud gennem årene, og hvordan har det udviklet sig? Det kan dette 20-årsjubilæum give anledning til at kigge nærmere på.

Internet og www

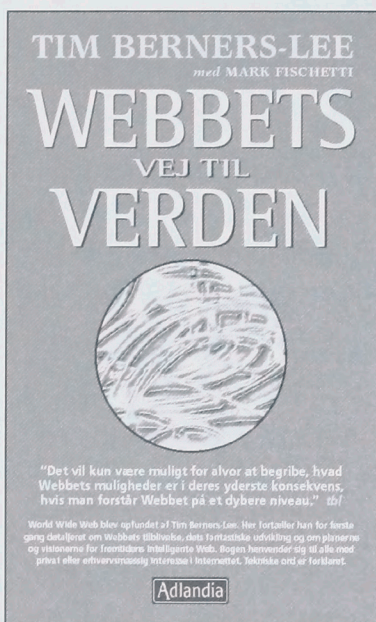
Der er imidlertid god grund til først kort at se på, hvad det egentlig var, Tim Berners-Lee opfandt. Hele wwvs forhistorie og

første leveår er levende beskrevet af Tim Berners-Lee selv i bogen *Webbets vej til verden*, og hans oprindelige post til nyhedsgruppen alt.hypertext kan stadig findes på nettet (se link på www.internethistorie.dk).

Når vi i dag går på nettet, mener vi i langt de fleste tilfælde 'the world wide web' eller bare 'web' eller 'www'. Men internet og www er ikke det samme. Internettet er en sammenkobling af computernetværk, men for at computerne i netværket kan kommunikere med hinanden, skal de have et sæt af fælles regler, det man med et teknisk ord kalder en protokol. Dem er der mange af, og www er blot en af dem. WWW er med andre ord en del af internettet, som benytter et bestemt sprog (http, html, url) for at få internettets computere til at finde hinanden, snakke sammen, og vise indholdet på en bestemt måde, nemlig som det vi kender som en webside.

WWW i 1991

I dag er www blevet så integreret en del af vores færden på nettet, at vi har svært ved overhovedet at se, hvorfor det egentlig var en rigtig god idé, dengang i 1991. Det særligt nyskabende ved Tim Berners-Lees opfindelse var, at han kombinerede to hidtil adskilte ting: dels den måde, man gennem 1980'erne havde brugt computerne på inden for det, man kalder hypertext, nemlig muligheden for at springe fra ét stykke tekst til et andet ved hjælp af et link, dels selve de netværksforbundne computere. Berners-Lee gjorde altså blot det, der i dag ser selvindlysende



*Tim Berners-Lee
Webbets vej
til verden (eng.
Weaving the
web) udkom på
dansk i 2001.
Tim Berners-Lee
er krediteret som
opfinder af world
wide web. I dag
er han professor
på Massachusetts
Institute of
Technology i
Boston, USA.*

ud, men som ikke var det i 1990'ernes begyndelse: han gjorde hyperteksten tilgængelig i computernetværkene, så man ikke bare brugte linket til at springe fra tekst til tekst på samme computer, men kunne springe til en tekst på en helt anden computer i netværket. Hyperlinket er altså en helt afgørende del af www.

En anden ny og væsentlig ting, som Berners-Lee lagde grundstenen til, var muligheden for at vise ikke blot skreven tekst, som på det tidspunkt stort set var den eneste måde, man kunne kommunikere på på internettet, men også billeder og lyd. WWW gjorde det muligt at kombinere skrift, billede og lyd på en meget nem måde, og med fremkomsten af de grafiske browsere, først Mosaic i 1993 og siden Netscape, gav det for alvor mening at tale om internettet som et multimedie. I begyndelsen gik det dog på grund af de små modemer relativt langsomt, og www blev hurtigt omdøbt til World Wide Wait, men i de år, der fulgte umiddelbart efter, blev www mere og mere grafisk i takt med at internethastigheden blev øget.

Det danske web

IDanmark gik opfindelsen af www relativt ubemærket hen i de første år, helt frem til 1993 var det andre systemer som gopher og WAIS, man satsede på, og for mange almindelige netbrugere var Bulletin Boards stadig synonym med nettet. Den første danske www-server etableredes som et eksperiment af Danmarks Tekniske Bibliotek i november 1992, og den første danske avis på nettet var *Ingeniøren* (december 1994).

De første år

Vi ved faktisk forbavsende lidt om, hvordan det danske web som et hele har set ud de første 10-12 år. Et væsentlig årsag hertil er den simple, at der ikke er foretaget systematisk indsamling og arkivering af det danske web før 2005. Og det, der ikke er indsamlet og gemt, er væk for stedse. Websider skal nemlig arkiveres, mens de stadig ligger på nettet – vi kan ikke vente med at gøre det til senere. Amerikanske undersøgelser fra omkring årtusindskiftet viser, at 40% af materialet på nettet forsvinder, inden der er gået et år, mens andre 40% er blevet ændret. Vi kan således kun forvente at finde 20% af det materiale, der var på nettet et år tidligere. Og over tid forsvinder det helt.

Vi mangler med andre ord en meget stor del af den danske kulturarv på www frem til 2005, hvor det danske internetarkiv Netarkivet bliver oprettet. Netarkivet er et samarbejde mellem Det Kongelige Bibliotek og Statsbiblioteket, og det varetager arkiveringen af det danske web, hvilket vil sige hele det danske

domæne .dk samt materiale på andre domæner (fx .com, .org, .nu, osv.), som er rettet mod en dansk offentlighed.

Der findes dog mindre samlinger af tidligt dansk web, dels i det nyåbnede webmuseum.dk, der dokumenterer og fortæller webbets historie (www.webmuseum.dk), dels i det amerikanske non-profit webarkiv The Internet Archive (www.archive.org), hvor man kan finde gamle udgaver af mange danske websites. Men der er ikke tale om systematiske indsamlinger med bevaring af kulturarven for øje.

The World Wide Web eller bare www er internetsoftware, der gør det muligt at se websider. WWW blev opfundet af den engelske computerspecialist Tim Berners-Lee, og den 6. august 1991 udsendte han den første udgave af programmet i flere nyhedsgrupper, til fri afbenyttelse og videreudvikling. Fra 1994 er det i hovedsagen organisationen W3C, der står for udvikling og regulering af wwws standarder.

Det danske web i Netarkivet

Det store problem er altså, at vi ikke har gemt det danske web for de første cirka 10 år. Men for perioden efter 2005 er vi meget bedre stillet.

Netarkivet har nu været i gang i seks år, og selvom det materiale, der findes i arkivet, af forskellige tekniske og praktiske grunde ikke kan siges at være helt komplet i forhold til, hvordan det danske web faktisk har set ud efter 2005, så er det så tæt på, som man overhovedet kan komme. Netarkivet indeholder i dag ca.

190TB og omtrent 5 milliarder digitale objekter.

Netarkivet er et samarbejde mellem Det Kongelige Bibliotek og Statsbiblioteket. Siden 2005 har Netarkivet indsamlet og gemt hele det danske domæne .dk samt materiale på andre domæner (fx .com, .org, .nu), som er rettet mod en dansk offentlighed. Netarkivets materiale er kun tilgængeligt for forskere.

Hvis man vil have et lille indblik i det danske webs udvikling de seneste fem år, kan der være god grund til at kigge i de enorme datamængder, der findes i Netarkivet. Til denne artikel har Netarkivet velvilligt udarbejdet tal, som gør det muligt at hæve sig op i helikopterperspektiv og se de store historiske linjer. Det er for første gang muligt at få et indblik i, hvordan hele det danske www har set ud og udviklet sig.

Konklusionen er den – måske lidt overraskende – at ganske almindelig skreven tekst er langt den mest fremherskende kommunikationsform på det danske web, og – måske mere overraskende – at det har den været stabilt i de seneste fem år, og faktisk med en svagt stigende tendens. Men lad os se lidt mere på, hvad Netarkivets tal kan fortælle os.

Når man skal tegne et billede af det danske web som et hele – og altså ikke kun se på enkelte websites og deres brug –, så er der ikke nogen entydig og gængs måde at gøre det på. Men et godt udgangspunkt vil under alle omstændigheder være at inddrage fire vigtige strøm-pile: mængden af digitalt materiale,

antallet af domæner med indhold, antallet af websites over en vis størrelse, og endelig antallet af de forskellige filtyper, fx billed- og tekstfiler.

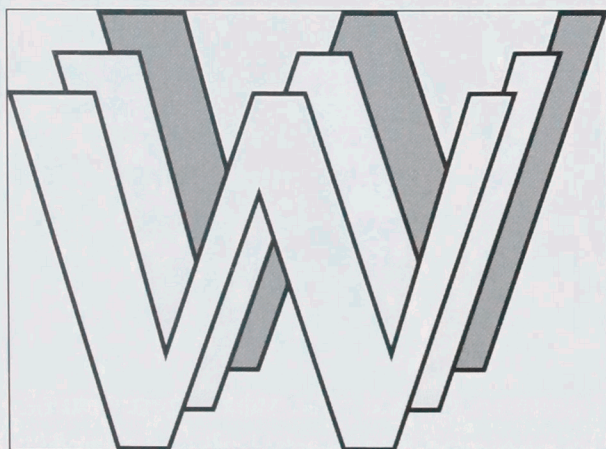
Det danske web fylder mere og mere

Der er ingen tvivl om, at det danske web er vokset fra 2005 og frem til i dag. Som nævnt arkiverer Netarkivet bl.a. hele det danske domæne .dk samt danske websites på andre domæner, og ser man på, hvor meget det fylder, så er det vokset fra omkring 8-10 TB i 2005/06 til omtrent 25TB i 2010, dvs. mere end en fordobling.

Mængden af domæner med indhold er ligeledes vokset, fra ca. 600.000 i 2006 til lige under 700.000 i 2007 og omtrent 900.000 i 2008, hvor domæneantallet topper og holder sig de følgende 2 år (ca. 20% af de registrerede danske domæner har ikke noget indhold).

Langt de fleste danske websites er under 10MB, dvs. de består ikke af meget mere end en forside samt ganske få undersider. Men ser man på antallet af websites som er større end 10MB,

Den engelske WWW's historiske logo, skabt af Robert Cailliau.



så er dette antal ligeledes vokset, fra ca. 50.000 i 2006 til ca. 120.000 i 2010, dvs. mere end en fordobling på under 5 år. Der er altså kommet flere større websites til.

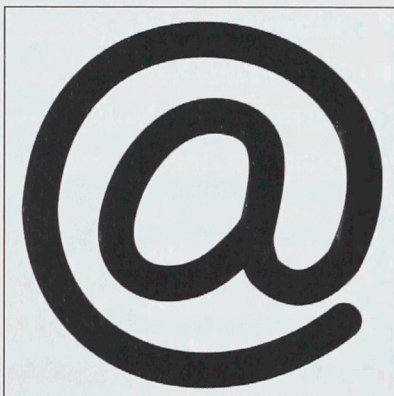
En webbruger, der satte sig til computeren i 2010 og begav sig ind på det danske web, ville altså have markant mere materiale, flere indholdsdomæner og flere store websites til rådighed, end hvis han satte sig til computeren i 2005.

Tekst er stadig meget fremherskende

Ser vi derimod på den sidste måde at tegne et billede af det danske web på – antallet af de enkelte filtyper som fx billed- eller tekstfiler – så er det slående, at til trods for at det samlede antal filer stiger til omtrent det dobbelte fra 2005 til 2010, så er den relative fordeling mellem filtyper stabil. I 2006 var 65% af de filer, der kom ind i Netarkivet, tekstfiler (typisk html-filer), mens 30% var billed- og grafikfiler (jpeg, .gif, png). Året efter ændredes disse tal til henholdsvis 74% og 22%, og der har de holdt sig bemærkelsesværdigt stabilt frem til i dag, og endog med en svag stigning i tekst-filer og tilsvarende fald i billedfiler i foråret 2011 (jf. grafik).

	Tekst	Billed/grafik	Andet
2006	65	30	5
2007	74	22	4
2008	71	23	6
2009	74	22	4
2010	73	21	6
2011 (forår)	77	17	6

Procentvis fordeling af filtyper i Netarkivet (kilde: Netarkivet, 2011)



Snabel-a er det danske navn for tegnet @, som bruges i email-adresser for at skille modtagerens navn fra domæne-adressen. Først benyttet af Ray Tomlinson i 1971.

Tallene dækker dog over en vis usikkerhed med hensyn til video, der af tekniske årsager ikke kan arkiveres fuldt ud, eksempelvis kan streamet video ikke arkiveres (det arkiverede video er indeholdt i kolonnen 'Andet'). Et forsigtigt skøn vil være, at antallet af videofiler ikke overstiger antallet af billedfiler, og at det procentvise antal tekstfiler derfor er tilsvarende mindre.

Men til trods for at tallene er baseret på, hvad der kom ind i Netarkivet, og at der er en vis usikkerhed omkring video, ligesom det heller ikke er alle tekst-filer, der faktisk svarer til tekst på skærmen, så er det ikke desto mindre slående, at det relative forhold mellem billeder/grafik og tekst på det danske web i de seneste 5 år er stort set uændret. Der er ganske vist kommet flere billeder, men der er også kommet mere skreven tekst. Det danske web er altså i hovedsagen meget stabilt baseret på det skrevne ord.

En webbruger, der i 2010 satte sig til sin computer og satte sig for at besøge samtlige danske websites, ville altså relativt set møde lige så meget tekst og lige så få billeder, som hvis han havde sat sig til rette i 2005.

Mange webbrugere har givetvis en erfaring af, at der er kommet flere og flere

billeder på web, men det skyldes nok, at de websites, hvor de fleste brugere færdes, også er dem, hvor der generelt er flest billeder, typisk nyhedssites og sociale netværkssteder som Facebook o.lign. Men forlader man det overordnede billede, som Netarkivets materiale tegner, så viser små stikprøver hurtigt, at den relative mængde af billeder på nyhedswebsites ikke er steget markant i perioden 2005-2010. Med andre ord: på de websites, hvor der altid har været billeder, er antallet af billeder per tekstside stadig omtrent det samme i dag som for 5 år siden.

Fremtidens danske web

Der er ingen tvivl om, at Tim Berners-Lee med skabelsen og udbredelsen af www i begyndelsen af 1990erne tog et væsentligt skridt i retning af at gøre internettet mere grafisk og dermed mere tilgængeligt for de fleste. Og springet fra det rent tekstbaserede til det grafiske net er så stort, at vi vanskeligt kan forestille os det i dag, hvor nettet bare er grafisk, med både skrift, billeder og video.

Men det ser altså også ud til, at skriftsproget stadig er den mest foretrukne kommunikationsform på det danske web, til trods for at øget båndbredde i har gjort det muligt at bruge flere billeder.

Det skal så blive interessant at se, om det danske web i de næste 20 år fortsætter med at være overvejende tekstbaseret, eller om de mange mobile medier med både billed- og videooptagere i stigende grad vil fungere som fødelinjer til et mere billedbaseret web.

Indkaldelse af ansøgninger til Frimærkehandler og jernbanehistoriker Peer Olav Thomassens Fond

Ansøgningsfrist 1. april 2012

Frimærkehandler og jernbanehistoriker Peer Olav Thomassens Fond

Fonden, som stiftet i 1992, har som formål „Videreførsel af forskningen i fortrinsvis dansk trafik- og kommunikationshistorie“ og yder støtte til

- a. anskaffelse af relevant materiale, herunder til den fornødne bevaring, registrering, katalogisering, indbinding, konservering mv. af det indgåede materiale
- b. udarbejdelse af publikationer, først og fremmest trykkomkostninger
- c. studie- og forskningsudgifter, herunder rejsetilskud, og
- d. tilskud til afholdelse af udstillinger, seminarer, kongresser mv.

Ad a: Støtte til opbygning af private bogsamlinger falder uden for fondens formål.

Ad c: Da fondens midler er af forholdsvis beskeden størrelse, vil det ikke være muligt at yde effektiv støtte til forskning inden for de i formålsparagraffen nævnte områder i form af løn.

Af hensyn til ekspedition af imødekomne ansøgninger, skal disse indeholde oplysning om modtagerens adresse og personnummer. Hvis ansøgeren er en juridisk person, skal man i ansøgningen oplyse CVR-nr.

Ansøgninger vedlagt curriculum vitae, eventuel publikationsliste samt anbefaling(er) stiles til Direktør Erland Kolding Nielsen og sendes til

Det Kongelige Bibliotek
Postboks 2149
1016 København K

TELEFON 33 47 47 47
TELEFAX 33 32 98 46

så den er biblioteket i hænde senest 1. april 2012 med morgenposten. Uddelingen forventes at finde sted maj/juni 2012.